

\*\*\*\*\*  
 W631631  
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
 Run on: Sat May 13 10:46:26 2000; Maspar time 4.61 Seconds  
 Tabular output not generated. 308.548 Million cell updates/sec

Title: >US-09-331-631-7  
 Description: (81-140) from US09331631.pep (3 of 3)  
 Perfect Score: 471  
 Sequence: 1 LQRYOQCGRCOEQOQCGOR.....HENYHNHKKRSEEEGQOR 60

Scoring table: PAM 150  
 Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: a.geneseq35  
 1:geneseqp

Statistics: Mean 24.413; Variance 96.857; scale 0.252

Prod. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	471	100.0	525	1 R62831	Theobroma cacao antimi	1.72e-38
2	471	100.0	566	1 R20181	Sequence encoded by 67	1.72e-38
3	181	38.4	590	1 W62832	Gossypium hirsutum ant	7.42e-09
4	177	37.6	666	1 W62829	Macadamia integrifolia	1.82e-08
5	170	36.1	666	1 W62828	Macadamia integrifolia	8.66e-08
6	163	34.6	625	1 W62830	Macadamia integrifolia	4.10e-07
7	117	24.8	28	1 W62841	Stenocarpus sinuatus a	8.49e-03
8	102	21.2	395	1 W03474	Mouse SRT-related prot	1.89e-01
9	100	21.2	637	1 W62837	Hordeum vulgare antimi	2.84e-01
10	99	21.0	521	1 W74802	Human secreted protein	3.47e-01
11	96	20.4	86	1 W95073	GST-HD fusion protein	6.36e-01
12	96	20.4	86	1 W95078	GST-HD fusion protein	6.36e-01
13	96	20.4	94	1 W95075	GST-HD fusion protein	6.36e-01
14	96	20.4	94	1 W95080	GST-HD fusion protein	6.36e-01
15	96	20.4	482	1 V07067	Renal cancer associate	6.36e-01
16	95	20.2	303	1 R60054	Dirofilaria immitis pa	7.78e-01
17	93	19.7	407	1 P94879	ORF incorporated withi	1.16e+00
18	93	19.7	444	1 W90340	G. max truncated SBP1	1.16e+00
19	93	19.7	524	1 W90339	G. max SBP1 protein.	1.16e+00
20	90	19.1	919	1 P93109	Human androgen recepto	2.10e+00
21	90	19.1	919	1 W14783	Androgen receptor.	2.10e+00
22	88	18.7	593	1 W62835	Zea mays antimicrobial	3.12e+00
23	87	18.5	96	1 W99838	HIV C765 protein seque	3.80e+00

24	87	18.5	409	1 W90342	G. max truncated SBP2	3.80e+00
25	87	18.5	489	1 W90341	G. max SBP2 protein.	3.80e+00
26	87	18.5	919	1 P90996	Human androgen recepto	3.80e+00
27	86	18.3	33	1 W62836	Zea mays antimicrobial	4.62e+00
28	86	18.3	96	1 W99831	HIV E21,24P protein se	4.62e+00
29	86	18.3	371	1 W73369	Epitope tagged TBP pro	4.62e+00
30	86	18.3	918	1 R12223	Human androgen recepto	4.62e+00
31	84	17.8	68	1 R75182	Partial peptide of hum	6.81e+00
32	84	17.8	96	1 W99836	HIV H71C protein seque	6.81e+00
33	84	17.8	131	1 R75181	Partial peptide of hum	6.81e+00
34	83	17.6	1206	1 R21079	Antimicrobial maize pe	8.27e+00
35	83	17.6	1206	1 W47030	Bovine N-proteinase.	8.27e+00
36	82	17.4	96	1 W99830	HIV Vpr protein sequen	1.00e+01
37	82	17.4	428	1 R73017	Murine U2AF heterodimer	1.00e+01
38	82	17.4	1529	1 R97985	CORR potassium channel	1.00e+01
39	81	17.2	96	1 W99833	HIV A30L protein seque	1.22e+01
40	80	17.0	69	1 R13399	Protein conty. Iga bin	1.47e+01
41	80	17.0	386	1 R04828	Arp 4.	1.47e+01
42	80	17.0	386	1 R06849	Protein Arp 4.	1.47e+01
43	80	17.0	404	1 W14909	Baker's yeast negative	1.47e+01
44	80	17.0	605	1 W62838	Glycine max antimicrob	1.47e+01
45	80	17.0	737	1 W18317	Drosophila Deltex prot	1.47e+01

## ALIGNMENTS

RESULT	1	ALIGNMENTS
ID	W62831 standard; Protein; 525 AA.	
AC	W62831;	
DE	27-OCT-1998 (first entry)	
DT	Theobroma cacao antimicrobial protein.	
OS	antimicrobial protein; infestation; control.	
PN	Theobroma cacao.	
PD	W09827805-AI.	
PF	02-JUL-1998.	
PR	22-DEC-1997; AU0874.	
PA	20-DEC-1996; AU-004275.	
PI	(RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.	
PI	Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;	
DR	WPI, 98-377279/32.	
PT	Novel anti-microbial protein from e.g. Macadamia integrifolia -	
PT	useful for controlling microbial infestations of plants or mammals	
PS	Claim 1; Page 47-49; 96pp; English.	
CC	The sequence is that of an antimicrobial protein which can	
CC	be used to control microbial infestations in plants and mammalian	
CC	animals.	
SQ	Sequence 525 AA:	
Query Match	100.0%; Score 471; DB 1; Length 525;	
Best Local Similarity	100.0%; Pred. No. 1.72e-38;	
Matches	60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
DB	81 LQRYOQCGRCOEQOQCGORCKWEYKEDRGCHENYHNHKKRSEEEGQOR 140	
QY	81 LQRYOQCGRCOEQOQCGOREQOQCKRKWEYKEDRGCHENYHNHKKRSEEEGQOR 140	
RESULT	2	
ID	R20181 standard; Protein; 566 AA.	
AC	R20181;	
DE	16-APR-1992 (first entry)	
DT	Sequence encoded by 67 kD T. cacao protein cDNA.	
KW	Cocoa; flavour; vicillin; seed storage protein.	
OS	Theobroma cacao.	
PN	W09119801-A.	
PD	26-DEC-1991.	
PF	07-JUN-1991; G00914.	
PR	11-JUN-1990; GB-013016.	
PA	(MRSC) MARS UK LTD.	
PI	Spencer ME, Hodge R, Deakin EA, Ashton S;	
DR	WPI; 92-024418/03.	
DR	N-PSDB; Q20377.	
PT	Recombinant cocoa proteins - are responsible for flavour in cocoa	



PR 11-APR-1997; US-043314.

PR 11-APR-1997; US-043314.





